baseline conditie (`conditie=1`) gelijk is aan $2.3021 - 3.0676 \times 1 + 0.3060 \times 0 = -0.7655$. Dit komt overeen met de geobserveerde steekproefgemiddeldes.

Voor de conditie met rugpijneducatie kunnen we afleiden dat het geschatte gemiddelde gelijk is aan $2.3021 - 3.0676 \times (-1) + 0.3060 \times (-1) = 5.0637$.

4 Toetsing

In dit stuk gaan we dieper in op hypothesetoetsen voor de parameters in een regressiemodel. Hierbij wordt ook de assumptie van normaal verdeelde fouttermen gemaakt:

$$\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

4.1 Modelvergelijkingen

Wanneer we meerdere (p) predictoren in het regressiemodel hebben, kunnen we ons afvragen of het nodig is die allemaal in het model op te nemen.

Om dit te toetsen kunnen we het model met p predictoren vergelijken met een model met k predictoren die een subset zijn van de p predictoren.

Dit is een speciaal geval van een algemene toets die nagaat of twee lineaire modellen, met het ene model genest in het andere, significant van elkaar verschillen wat betreft de mogelijkheid om de uitkomst Y te voorspellen.

Een lineair model B is genest in model A wanneer bij model B bijkomende lineaire restricties m.b.t. de te schatten parameters opgelegd worden.

Als we de 2 modellen A (zonder restricties) en B (met restricties) met elkaar wensen te vergelijken, kunnen we gebruik maken van deze toetsingsgrootheid:

$$F = \frac{(\mathrm{SSE}_B - \mathrm{SSE}_A)/(\mathrm{df}_B - \mathrm{df}_A)}{\mathrm{SSE}_A/\mathrm{df}_A} \tag{2}$$

SSE_A stelt de fout kwadratensom van model A voor en SSE_B de fout kwadratensom van model B. df_A en df_B stellen de overeenkomstige vrijheidsgraden in respectievelijk model A en model B voor.

Met p predictoren in het model dienen op basis van n observaties $p + 1$ coëfficiënten geschat te worden (vergeet het intercept niet). Dit betekent dat in dit geval $\mathrm{df}_A = n - (p + 1)$.

Wanneer model B slechts een subset van k predictoren bevat, is $\mathrm{df}_B = n - (k + 1)$.

Er kan aangetoond worden dat onder H_0 (beide modellen zijn niet verschillend van elkaar), de toetsingsgrootheid (2) F-verdeeld is met $\text{df}_B - \text{df}_A$ vrijheidsgraden voor de teller en df_A vrijheidsgraden voor de noemer ($F \sim F(\text{df}_B - \text{df}_A, \text{df}_A)$). Dit is de nuldistributie van F in (2), dit betekent dat we de bijhorende p-waarde van de toets kunnen berekenen en beslissen om H_0 al dan niet te verwerpen.

Dit is een algemene beschrijving voor toetsing aan de hand van modelvergelijkingen. In de volgende secties worden enkele concrete gevallen in meer detail besproken.

4.2 Toets voor alle predictoren

We stellen ons hier de vraag: is er tenminste 1 predictor nuttig in het voorspellen van de uitkomst?

We vergelijken het model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1, \ldots, n \tag{3}$$

met het model zonder predictoren:

$$Y_i = \beta_0 + \varepsilon_i.$$

Dit model noemen we het **nulmodel**. Het aantal vrijheidsgraden dat geassocieerd is met het nulmodel is $n - 1$ (er dient enkel een intercept geschat te worden). Voor het nulmodel geldt dat $\hat{Y}_i = B_0 = \bar{Y}$ ($i = 1, \ldots, n$).

Of nog: we toetsen de volgende nulhypothese

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$$

tegenover de alternatieve hypothese H_1 die stelt dat tenminste 1 regressiecoëfficiënt verschillend is van 0. Het verwerpen van H_0 betekent dus dat we verder moeten nagaan voor welke predictoren de regressiecoëfficiënt significant verschillend is van 0.

Er geldt dat de fout kwadratensom voor het nulmodel gelijk is aan de totale kwadratensom SST van het volledig model in (3). Het aantal vrijheidsgraden dat overeenkomt met SST is $n - 1$.

Bijgevolg ziet de toetsingsgrootheid in (2) er in dit geval als volgt uit:

$$F = \frac{(\text{SST} - \text{SSE})/(\text{df}_0 - \text{df}_A)}{\text{SSE}/\text{df}_A} = \frac{\text{SSR}/(\text{df}_0 - \text{df}_A)}{\text{SSE}/\text{df}_A}$$

met df_0 het aantal vrijheidsgraden van het nulmodel ($= n - 1$) en df_A het aantal vrijheidsgraden van het volledige model in (3) ($= n - (p + 1)$). Bijgevolg is $\text{df}_0 - \text{df}_A = p$. Dit is het aantal vrijheidsgraden dat hoort bij de verklaarde kwadratensom SSR van een model en is gelijk aan het aantal parameters van het volledige model, intercept niet meegerekend.

Beschouw het voorbeeld rond overclaiming (zie sectie 1.2.1). We voorspellen overclaiming op basis van gepercipieerde kennis, accuraatheid en FINRA-score.

```
> fit3_expertise<-lm(overclaiming_proportion~self_perceived_knowledge+accuracy
             +FINRA_score,data=expertise)
> summary(fit3_expertise)

Call:
lm(formula = overclaiming_proportion ~ self_perceived_knowledge + accuracy
         + FINRA_score, data = expertise)

Residuals:
    Min       1Q   Median       3Q      Max
-0.38033 -0.08672 -0.01418  0.08808  0.30886

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.057787   0.039203   1.474   0.1421
self_perceived_knowledge  0.094069   0.008018  11.732   <2e-16 ***
accuracy                 -0.793219   0.045655 -17.374   <2e-16 ***
FINRA_score               0.018370   0.008576   2.142   0.0334 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.1256 on 198 degrees of freedom
Multiple R-squared:  0.7116,Adjusted R-squared:  0.7073
F-statistic: 162.9 on 3 and 198 DF,  p-value: < 2.2e-16
```

In deze output kunnen we op de onderste lijn het resultaat van de F-toets voor alle predictoren aflezen. Hier is de geobserveerde toetsingsgrootheid f^* gelijk aan 162.9, bij de bijhorende p-waarde duidt `< 2.2e-16` aan dat deze p-waarde heel klein is. We hebben sterke evidentie tegen de nulhypothese wat impliceert dat we evidentie hebben voor het feit dat minstens 1 predictor een invloed heeft op de uitkomst. In dit voorbeeld is het aantal observaties $n = 202$. De toetsingsgrootheid volgt onder de nulhypothese een F-verdeling met 3 vrijheidsgraden (aantal parameters dat getoetst wordt) voor de teller en 198 (i.e. 202-(3+1)) voor de noemer.

In dit stukje van de output kunnen we de kwadratensommen zelf niet aflezen. Deze informatie krijgen we wel als we zelf de modelvergelijkingstoets uitvoeren via het commando `anova`.

Het nulmodel kunnen we als volgt definiëren:

```
fit0_expertise<-lm(overclaiming_proportion~1,data=expertise)
```

Het resultaat van de modelvergelijking die het effect voor alle predictoren toetst is als volgt:

```
> anova(fit0_expertise,fit3_expertise)
Analysis of Variance Table

Model 1: overclaiming_proportion ~ 1
Model 2: overclaiming_proportion ~ self_perceived_knowledge + accuracy +
FINRA_score
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    201 10.840
2    198  3.126  3    7.7139 162.87 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Bovenaan in de output staat weergegeven welke 2 modellen met elkaar vergeleken worden.

In de kolom `RSS` lezen we de fout kwadratensommen af. Voor het nulmodel is deze som gelijk aan 10.840. Dit is ook de SST van het volledige model. Het overeenkomstige aantal vrijheidsgraden kan afgelezen worden in de kolom `Res.Df` en is gelijk aan $n - 1 = 202 - 1 = 201$.

Voor het volledige model is de fout kwadratensom SSE gelijk aan 3.126, het overeenkomstig aantal vrijheidsgraden is gelijk aan $n - (3 + 1) = 202 - 4 = 198$. We lezen op de onderste lijn onder `Sum of Sq` verder af dat SST-SSE=7.7139=SSR en dat het overeenkomstig aantal vrijheidsgraden gelijk is aan 3, i.e. het verschil in aantal vrijheidsgraden tussen het volledige model en het nulmodel of nog, het aantal parameters voor de predictoren in het volledige model, intercept niet inbegrepen.

We zien dat de geobserveerde toetsingsgrootheid inderdaad overeenkomt met wat we voordien bekwamen, namelijk $f^* = (7.7139/3)/(3.126/198) = 162.87$.

4.3 Toets voor een subset van predictoren

De toets voor alle predictoren uit de vorige sectie is een speciaal geval van de modelvergelijkingstoets waarbij een set van predictoren getoetst wordt. Wanneer we in het voorbeeld rond overclaiming (zie sectie 1.2.1) wensen te toetsen of het model waar naast de gepercipieerde kennis ook accuraatheid en FINRA-score als predictoren opgenomen zijn de variatie in overclaiming beter verklaart, vergelijken we het model met de 3 predictoren met het model met enkel gepercipieerde kennis als predictor.

Als β_2 en β_3 de regressiecoëfficiënten voor respectievelijk accuraatheid en FINRA-score voorstellen, dan toetsen we $H_0 : \beta_2 = \beta_3 = 0$ versus de alternatieve hypothese die stelt dat minstens 1 van beide parameters niet 0 is. De specifiëring in R van de 2 modellen gebeurt als volgt:

```
fit1_expertise<-lm(overclaiming_proportion~self_perceived_knowledge,data=expertise)
fit3_expertise<-lm(overclaiming_proportion~self_perceived_knowledge+accuracy
                    +FINRA_score,data=expertise)
```

Wanneer we beide modellen met elkaar vergelijken, krijgen we:

```
> anova(fit1_expertise,fit3_expertise)
Analysis of Variance Table

Model 1: overclaiming_proportion ~ self_perceived_knowledge
Model 2: overclaiming_proportion ~ self_perceived_knowledge + accuracy +
FINRA_score
   Res.Df  RSS Df   Sum of Sq    F    Pr(>F)
1    200 8.3303
2    198 3.1260  2    5.2044 164.82 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Model 1 is hier genest in model 2, i.e. in de notatie in (2) is model 1 dus gelijk aan model B terwijl model 2 gelijk is aan model A.

Het aantal vrijheidsgraden (`Res.Df`) voor model 1 is 200 (df_B), i.e. 202-(1+1) terwijl dit 198 (df_A) voor model 2 is.

In kolom `RSS` zien we dat $\mathrm{SSE}_B = 8.3303$ en $\mathrm{SSE}_A = 3.1260$.

Op de onderste lijn zien we verder dat er $200 - 198 = 2 = \mathrm{df}_B - \mathrm{df}_A$ parameters getoetst worden.

De geobserveerde toetsingsgrootheid is $f^* = (5.2044/2)/(3.1260/198) = 164.82$. Onder H_0 is deze toetsingsgrootheid F-verdeeld met 2 en 198 vrijheidsgraden. De bijhorende p-waarde (`Pr(>F)`) is heel klein waardoor we evidentie hebben tegen de nulhypothese en aannemen dat model 2 de variatie in overclaiming beter verklaart dan model 1.

4.4 Toets voor 1 predictor

We stellen ons hier de vraag: kan 1 welbepaalde predictor uit het model weggelaten worden?

Hiervoor kunnen we opnieuw een modelvergelijkingstoets uitvoeren: het volledige model wordt vergeleken met het model zonder de predictor.

Als deze predictor van intervalniveau is, komt dit overeen met het toetsen van 1 enkele parameter.

Als de predictor van nominaal niveau is, zal het aantal parameters dat getoetst wordt overeenkomen met het aantal hulpveranderlijken voor deze predictor. M.a.w. als we willen nagaan of een nominale predictor een invloed heeft op de uitkomst, vergelijken we de modellen met en zonder de hulpveranderlijken die coderen voor de nominale predictor. Het resultaat van deze toets zal onafhankelijk zijn van het coderingsschema (dummy- of effect-codering) dat gebruikt wordt voor de nominale predictor.

4.4.1 Predictor van intervalniveau

Beschouw het volgende model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_\ell x_{i\ell} + \ldots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1, \ldots, n$$

Wanneer we wensen te toetsen of predictor ℓ uit het model weggelaten kan worden, toetsen we $H_0 : \beta_\ell = 0$ tegenover $H_1 : \beta_\ell \neq 0$.

In het geval dat er slechts 1 parameter getoetst wordt, is de F-toets uit de modelvergelijking equivalent aan een t-toets. Hiervoor kunnen we gebruik maken van de volgende toetsingsgrootheid:

$$T = \frac{B_\ell}{S_{B_\ell}}$$

waarbij S_{B_ℓ} de standaardfout van B_ℓ voorstelt. Er kan aangetoond worden dat onder H_0, $T \sim t_{n-(p+1)}$ of nog: dat T onder H_0 een t-verdeling volgt met $n - (p + 1)$ vrijheidsgraden.

Onder H_0 is de absolute waarde van de toetsingsgrootheid klein (dicht bij 0). Als H_1 geldt, zal de absolute waarde groot zijn.

Op basis van de geobserveerde toetsingsgrootheid $t^* = b_\ell / s_{B_\ell}$ kunnen we de p-waarde berekenen:

$$2 \times P\left(T \geq \left|\frac{b_\ell}{s_{B_\ell}}\right|\right) \text{ met } T \sim t_{n-(p+1)}.$$

Via de output van `lm` in R krijgen we standaard het resultaat van de t-toets voor alle parameters in het model (ook voor het intercept, maar hieraan wordt in de praktijk doorgaans geen aandacht aan besteed).

Herneem het voorbeeld rond overclaiming (zie sectie 1.2.1).

```
> fit3_expertise<-lm(overclaiming_proportion~self_perceived_knowledge+accuracy
+FINRA_score,data=expertise)
> summary(fit3_expertise)

Call:
lm(formula = overclaiming_proportion ~ self_perceived_knowledge + accuracy
+ FINRA_score, data = expertise)

Residuals:
Min       1Q   Median       3Q      Max
-0.38033 -0.08672 -0.01418  0.08808  0.30886

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.057787   0.039203   1.474   0.1421
self_perceived_knowledge  0.094069   0.008018  11.732   <2e-16 ***
accuracy                 -0.793219   0.045655 -17.374   <2e-16 ***
FINRA_score               0.018370   0.008576   2.142   0.0334 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.1256 on 198 degrees of freedom
Multiple R-squared:  0.7116,Adjusted R-squared:  0.7073
F-statistic: 162.9 on 3 and 198 DF,  p-value: < 2.2e-16
```

Voor iedere predictor lezen we in de kolom `t value` de geobserveerde toetsingsgrootheid af en in de kolom `Pr(>|t|)` de overeenkomstige p-waarde.

Zo zien we bvb. voor FINRA-score dat de geobserveerde toetsingsgrootheid $t^* = 0.018370/0.008576 = 2.142$. De bijhorende p-waarde is 0.0334, dit is kleiner dan 5% wat betekent dat we de nulhypothese dat FINRA-score geen effect heeft op overclaiming kunnen verwerpen op het 5% significantieniveau.

Via `anova` waarbij we de modelvergelijking zelf definiëren komen we tot hetzelfde resultaat:

```
fit2_expertise<-lm(overclaiming_proportion~self_perceived_knowledge+accuracy,
                 data=expertise)
fit3_expertise<-lm(overclaiming_proportion~self_perceived_knowledge+accuracy
                +FINRA_score,data=expertise)
> anova(fit2_expertise,fit3_expertise)
Analysis of Variance Table

Model 1: overclaiming_proportion ~ self_perceived_knowledge + accuracy
Model 2: overclaiming_proportion ~ self_perceived_knowledge + accuracy +
FINRA_score
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1    199 3.1984
2    198 3.1260  1  0.072431 4.5878 0.03342 *
```

Er geldt dat $f^* = t^{*2} = (2.142)^2 = 4.59$.

4.4.2 Predictor van nominaal niveau

Om het effect van een nominale predictor met I niveaus te toetsen, voeren we een modelvergelijkingstoets uit met als nulhypothese dat de regressiecoëfficiënten die horen bij de $(I-1)$ hulpveranderlijken allen 0 zijn.

Dit komt neer op toetsen of de gemiddelde uitkomst gelijk is over de verschillende niveaus (conditioneel op de overige predictoren in het model).

Merk op dat wanneer $I = 2$ er slechts 1 parameter getoetst moet worden. In dat geval is de F-toets ook equivalent aan de t-toets zoals in voorgaande sectie.

We hernemen het voorbeeld van de pijneducatie (sectie 1.2.2) waarbij we het effect van `conditie` (3 niveaus) op de uitkomst `buig` regresseren. De output bij dummy-codering is als volgt:

```
> fit_pijneducatie_dummy<-lm(Buig~Conditie,data=pijneducatie)
```

```
> summary(fit_pijneducatie_dummy)

Call:
lm(formula = Buig ~ Conditie, data = pijneducatie)

Residuals:
Min       1Q   Median       3Q      Max
-25.3164  -5.5974  -0.0914   4.7621  26.7199

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -0.7655     1.5885  -0.482   0.6308
ConditieAlgemene pijneducatie   3.3736     2.2464   1.502   0.1358
ConditieRugpijneducatie         5.8291     2.2327   2.611   0.0102 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 10.05 on 118 degrees of freedom
Multiple R-squared:  0.05497,Adjusted R-squared:  0.03895
F-statistic: 3.432 on 2 and 118 DF,  p-value: 0.03559
```

Aangezien `conditie` de enige predictor in het model is, lezen we hier onmiddellijk het resultaat voor de modelvergelijkingstoets voor het effect van `conditie` af (door vergelijking van het model met nulmodel): de geobserveerde toetsingsgrootheid f^* is gelijk aan 3.432 met een bijhorende p-waarde gelijk aan 0.036. We kunnen bijgevolg de nulhypothese die stelt dat de gemiddelde verschilscore in voorover buigen gelijk is in de 3 condities, verwerpen.

Dit resultaat kunnen we ook als volgt bekomen:

```
> fit0_pijneducatie_dummy<-lm(Buig~1,data=pijneducatie)
> anova(fit0_pijneducatie_dummy,fit_pijneducatie_dummy)
Analysis of Variance Table

Model 1: Buig ~ 1
Model 2: Buig ~ Conditie
Res.Df   RSS     Df Sum of Sq   F  Pr(>F)
1    120 12602
2    118 11910  2    692.76 3.4319 0.03559 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Kijk zelf na dat het resultaat hetzelfde blijft wanneer effect-codering gebruikt wordt voor `conditie`.

4.4.3 Algemene strategie: Anova-tabel in R

Van zodra er meerdere predictoren (waaronder nominale predictoren) in een regressiemodel opgenomen zijn, is het praktischer om de resultaten voor de toetsen van alle predictoren afzonderlijk via 1 enkel commando te verkrijgen. Het commando `Anova` (let op de hoofdletter!) uit het R-package `car` geeft deze resultaten in 1 anova-tabel weer. De reden dat we werken met dit commando is dat het toelaat verschillende types toetsen uit voeren waaronder deze die overeenkomen met wat andere software zoals SPSS standaard zou weergeven.

Wij maken altijd gebruik van Type III-toetsen bij het toetsen van de predictoren afzonderlijk. Het onderscheid tussen de verschillende types toetsen speelt voornamelijk een rol wanneer ook interacties in het model opgenomen zijn. We komen er op terug in sectie 5 over interacties.

Een belangrijke opmerking bij het gebruik van de Type III-toetsen bij Anova is dat we voor het toetsen zelf eerst moeten overgaan op effect-codering van de nominale variabelen in het model. Onthoud dat een model met effect-codering in essentie hetzelfde is als een model met dummy-codering; enkel de interpretatie van de parameters wijzigt. Het betreft dus louter een technisch aspect om correcte en zinvolle resultaten bij het toetsen te krijgen. Dit is opnieuw enkel van belang als er ook interacties met nominale predictoren opgenomen zijn (zie verder).

We hernemen het voorbeeld rond pijneducatie (sectie 1.2.2) . In deze studie zijn de participanten niet at random toegewezen aan de verschillende condities. Daarom is het zinvol om te corrigeren voor het effect van leeftijd en de graad van depressie. Leeftijd (`Leeft`) en Depressiescore (`Dep`) mogen als van intervalniveau verondersteld worden.

```
> fit3_pijneducatie_dummy<-lm(Buig~Conditie+Leeft+Dep,data=pijneducatie)
> summary(fit3_pijneducatie_dummy)

Call:
lm(formula = Buig ~ Conditie + Leeft + Dep, data = pijneducatie)

Residuals:
Min       1Q   Median       3Q      Max
-20.2287  -4.9537  -0.5254   5.1798  18.7101

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)                  16.42433    4.41878   3.717 0.000312 ***
ConditieAlgemene pijneducatie  3.73261    1.72868   2.159 0.032891 *
ConditieRugpijneducatie        4.78874    1.72026   2.784 0.006276 **
Leeft                         -0.10238    0.10749  -0.952 0.342840
Dep                           -0.65121    0.07187  -9.061 3.75e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.723 on 116 degrees of freedom
Multiple R-squared:  0.451,Adjusted R-squared:  0.4321
F-statistic: 23.82 on 4 and 116 DF,  p-value: 2.133e-14
```

Uit bovenstaande output kunnen we aflezen dat leeftijd in het model geen statistisch significante invloed heeft op de uitkomst ($p = 0.34$) en depressiescore wel ($p < 0.001$), maar voor conditie kunnen we dit niet rechtstreeks aflezen. Wat we aflezen is het resultaat van de toetsen waarbij de conditie met algemene pijneducatie en rugpijneducatie afzonderlijk met de baseline conditie vergeleken worden.

We vragen nu de bijhorende anova-tabel op om de resultaten voor alle predictoren afzonderlijk te zien. Hiervoor herdefiniëren we eerst het model aan de hand van effect-codering.

```
> library(car)
> fit3_pijneducatie_test<-lm(Buig~Conditie+Leeft+Dep,data=pijneducatie,
                                  contrasts=list(Conditie=contr.sum))
> Anova(fit3_pijneducatie_test,type=3)
Anova Table (Type III tests)

Response: Buig
           Sum Sq  Df F value    Pr(>F)
(Intercept) 1208.3   1 20.2587 1.618e-05 ***
Conditie     509.2   2  4.2685   0.01626 *
Leeft         54.1   1  0.9072   0.34284
Dep         4897.2   1 82.1057 3.746e-15 ***
Residuals   6918.8 116
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

In deze output lezen we voor iedere predictor het resultaat af van de modelvergelijkingstoets waarbij de modellen met en zonder de predictor vergeleken worden (de geobserveerde toetsingsgrootheid in `F value` en p-waarde in `Pr(>F)`). We krijgen de overeenkomstige kwadratensom voor deze toetsen (`Sum Sq`) en het aantal vrijheidsgraden (`Df`), i.e. het aantal parameters dat getoetst wordt (1 voor leeftijd en 1 voor depressiescore maar 2 voor conditie).

Voor `Leeft` en `Dep` bekomen we (uiteraard) dezelfde resultaten voor de F-toets als voor de t-toets. We lezen verder af dat het effect van conditie na correctie voor leeftijd en depressiescore nog steeds statistisch significant is op het 5% significantieniveau ($p = 0.016$).

Verifieer zelf dat via `Anova` dezelfde resultaten bekomen worden voor de analyses uitgevoerd in sectie 4.4.1 en 4.4.2!

5 Interactie (moderatie)

5.1 Wat is interactie?

Bij een regressiemodel met 2 predictoren X_1 en X_2 waarvoor

$$E(Y_i|x_{i1}, x_{i2}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \ \ (i = 1, \ldots, n)$$

zijn de effecten van de predictoren additief: het effect van de ene predictor hangt niet af van het niveau van de andere, hun gezamenlijk effect kan gemodelleerd worden als een som.

Wanneer er een **interactie** is tussen beide predictoren betekent dit dat het effect van een combinatie van de 2 predictoren groter of kleiner is dan de som van de afzonderlijke effecten. Het effect van de ene predictor is nu anders voor elk niveau van de andere predictor. We hebben dan:

$$E(Y_i|x_{i1}, x_{i2}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} \ \ (i = 1, \ldots, n)$$

De interactieterm is gelijk aan het product van beide predictoren. β_3 is het interactie-effect.

De termen **effect-modificatie** en **moderatie** worden ook vaak gebruikt om te verwijzen naar een interactie. X_1 modificeert het effect van X_2 op Y (en vice versa). X_1 is een moderator voor het effect van X_2 op Y (en vice versa).

De regressiecoëfficiënten β_1 en β_2 hebben nu een andere interpretatie dan voorheen. Wanneer X_1 1 eenheid stijgt terwijl $X_2 = x_2$ constant blijft, neemt de verwachte uitkomst toe met $\beta_1 + \beta_3 x_2$.

Analoog: wanneer X_2 1 eenheid stijgt terwijl $X_1 = x_1$ constant blijft, neemt de verwachte uitkomst toe met $\beta_2 + \beta_3 x_1$.